CHAPTER 8 - Limit Theorems

8.1 - Introduction

The most important theoretical result in probability theory are limit theorems.

↳ 2 of these: laws of large numbers, central limit theorem *

- Theorems are considered to be laws of large numbers if they are concerned with stating conditions under which the average of a sequence of a random variables converges to the expected average
- By contrast, central limit theorems are concerned with determining conditions under which the sum of a

large number of random variables has a probability distribution that is approximately normal.

8.2 - CHEBYSHEV'S INEQUALITY and the WEAK LAW OF LARGE NUMBERS

- We start this section by proving a result known as Markov's inequality

Proposition 2.1. Markov's inequality
If X is a random variable that takes only nonnegative values, then, for any v[alue]
$a > 0$,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proof. For $a > 0$, let

$$I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$$

and note that, since $X \geq 0$,

$$I \leq \frac{X}{a}$$

Taking expectations of the preceding inequality yields

$$E[I] \leq \frac{E[X]}{a}$$

which, because $E[I] = P\{X \geq a\}$, proves the result.

As a corollary, we obtain Proposition 2.2.

Proposition 2.2. Chebyshev's inequality
If X is a random variable with finite mean μ and variance σ^2, then, for any value $k > 0$,

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Proof. Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E[(X - \mu)^2]}{k^2} \tag{2.1}$$

But since $(X - \mu)^2 \geq k^2$ if and only if $|X - \mu| \geq k$, Equation (2.1) is equivalent to

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

and the proof is complete. □

* The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known.

EXAMPLE 2a

Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

a) What can be said about the probability that this week's production will exceed 75?

→ Let X be the number of items that will be produced in a week

Markov's inequality:

$$P(X > 75) = \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 & 60.?

→ By Chebyshev's inequality.

$$P(|X-50| \geq 10) \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

Hence:

$$P(|X-50| < 10) \geq 1-\frac{1}{4} = \frac{3}{4}$$

So the probability that this week's production will be between 40 & 60 is at least 75%

- As Chebyshev's inequality is valid for all distributions of the random variable X, we cannot expect the bound on the probability to be very close to the actual probability in most cases

EXAMPLE 2b

If $X \sim U(0,10)$, then $E[X] = 5$ &

$Var(X) = \frac{25}{3}$

it follows from Chebyshev's inequality that,

$$P(|X-5|>4) \le \frac{25}{3(16)} \simeq 0.52$$

whereas the exact result is

$P(|x-5|>4) = 0.20$

Although Chebyshev's inequality is correct the upper bound that it provides is not particularly close to the actual probability.

Similarly, if X is a normal random variable with mean μ & variance σ^2 Chebyshev's inequality states that

$$P(|X-\mu| > 2\sigma) \leq \frac{1}{4}$$

whereas the actual probability is given

$$P\{|X-\mu| > 2\sigma\} = P\left\{\left|\frac{X-\mu}{\sigma}\right| > 2\right\} = 2[1-\Phi(2)] \approx .0456$$

Note:

Chebyshev's inequality is often used as a theoretical tool in proving results. ↳ This is illustrated. First by Proposition 2.3 and then, most importantly by the weak-law of large numbers

Proposition 2.3. If $\text{Var}(X) = 0$, then

$$P\{X = E[X]\} = 1$$

In other words, the only random variables having variances equal to 0 are those which are constant with probability 1.

Theorem 2.1 The weak law of large numbers
Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \quad \text{as} \quad n \to \infty$$

Proof. We shall prove the theorem only under the additional assumption that the random variables have a finite variance σ^2. Now, since

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \quad \text{and} \quad \text{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

it follows from Chebyshev's inequality that

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

and the result is proven. □

8.3 – The Central Limit Theorem

- In simple words:

↳ It states that the sum of a large number of independent random variables has a distribution that is approximately normal. Hence, it not only provides a simple method for computing approximate probabilities for sums of independent random variables, but also helps explain the remarkable fact that the empirical frequencies of so many natural populations exhibit bell-shaped (normal) curves.

Theorem 3.1 The central limit theorem

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx \quad \text{as} \quad n \to \infty$$

The key to the proof of the central limit theorem is the following lemma, which we state without proof.

{ Proof of CLT : pg 407 - 408 }

EXAMPLE 3a :

An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to be reasonably sure that his estimated distance is accurate to within $\pm .5$ light-year?

(see pg 409 - 411) solution

If $X_1, \ldots, X_n$ are the n-measurement →

$$Z_n = \frac{\sum_{i=1}^{n} X_i - nd}{2\sqrt{n}}$$

8.4 - The Strong Law of Large Numbers

↳ This is probably the best known result in probability theory.

→ It states that the average of a sequence of independent random variables having a common distribution will, with probability 1, converge to the mean of that distribution.

Theorem 4.1 The strong law of large numbers
Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \qquad \text{as} \qquad n \to \infty^{\dagger}$$

As an application of the strong law of large numbers, suppose that a sequence of independent trials of some experiment is performed. Let E be a fixed event of the experiment, and denote by $P(E)$ the probability that E occurs on any particular trial. Letting

$$X_i = \begin{cases} 1 & \text{if } E \text{ occurs on the } i\text{th trial} \\ 0 & \text{if } E \text{ does not occur on the } i\text{th trial} \end{cases}$$

we have, by the strong law of large numbers, that with probability 1,

$$\frac{X_1 + \cdots + X_n}{n} \to E[X] = P(E) \tag{4.1}$$

Since $X_1 + \cdots + X_n$ represents the number of times that the event E occurs in the first n trials, we may interpret Equation (4.1) as stating that, with probability 1, the limiting proportion of time that the event E occurs is just $P(E)$.

Although the theorem can be proven without this assumption, our proof of the strong law of large numbers will assume that the random variables X_i have a finite fourth moment. That is, we will suppose that $E[X_i^4] = K < \infty$.

(Proof can be found pg 415)

8-5 - OTHER INEQUALITIES

→ We are sometimes conforonted with situations in which we are interested in obtaining an upper bound for a probability of the form $P(X - \mu \geq a)$, where a is some positive value and when only the mean $\mu = E[X]$ & variance $\sigma^2 = Var(X)$ of the distribution of X are known.

- Of course, since $X - \mu \geq a > 0$ implies that $|X - \mu| \geq a$

→ Chebyshev's inequality

$$P\{X - \mu \geq a\} \leq P\{|X - \mu| \geq a\} \leq \frac{\sigma^2}{a^2} \quad \text{when} \quad a > 0$$

→ Check out Proposition 5.1 pg 418

Proposition 5.2 pg 422

Proposition 5.3 pg 424

SUMMARY

SUMMARY

Two useful probability bounds are provided by the *Markov* and *Chebyshev* inequalities. The Markov inequality is concerned with nonnegative random variables and says that, for X of that type,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

for every positive value a. The Chebyshev inequality, which is a simple consequence of the Markov inequality, states that if X has mean μ and variance σ^2, then, for every positive k,

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

The two most important theoretical results in probability are the *central limit theorem* and the *strong law of large numbers*. Both are concerned with a sequence of independent and identically distributed random variables. The central limit theorem says that if the random variables have a finite mean μ and a finite variance σ^2, then the distribution of the sum of the first n of them is, for large n, approximately that of a normal random variable with mean $n\mu$ and variance $n\sigma^2$. That is, if $X_i, i \geq 1$, is the sequence, then the central limit theorem states that, for every real number a,

$$\lim_{n \to \infty} P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx$$

The *strong law of large numbers* requires only that the random variables in the sequence have a finite mean μ. It states that, with probability 1, the average of the first n of them will converge to μ as n goes to infinity. This implies that if A is any specified event of an experiment for which independent replications are performed, then the limiting proportion of experiments whose outcomes are in A will, with probability 1, equal $P(A)$. Therefore, if we accept the interpretation that "with probability 1" means "with certainty," we obtain the theoretical justification for the long-run relative frequency interpretation of probabilities.